

# Prevenção recém-nascidos com baixo peso - Uma abordagem de Aprendizado de Máquinas utilizando o DATASUS

Gabriel A. Miranda

<sup>1</sup>Instituto de Computação – Universidade Federal Fluminense (UFF)  
Niterói – RJ – Brazil

**Abstract.** *The birth weight of a newborn is an important factor for their health and development. An individual born with a weight below the expected can pose risks both to themselves and to the mother. Low birth weight is associated with language development issues. Being able to predict the risk of low birth weight can be an important medical tool. Previous studies with the same objective have used both maternal and neonatal information. This study proposes developing a Machine Learning model to predict the risk of low birth weight based on demographic data of the mother and father, available in the SINASC DATASUS database.*

**Resumo.** *O peso do recém-nascido é um fator importante para sua saúde e seu desenvolvimento. Um indivíduo que nasce com peso abaixo do esperado pode representar riscos tanto para si quanto para a mãe. Baixo peso ao nascer está relacionado a problemas no desenvolvimento da linguagem. Conseguir prever o risco de baixo peso pode ser uma ferramenta médica importante. Trabalhos anteriores com o mesmo objetivo utilizaram tanto informações da mãe como informações do bebê, a partir de exames neonatal. Este trabalho propõe desenvolver um modelo de Machine Learning para prever o o risco de baixo peso ao nascer a partir de dados demográficos da mãe e do pai, disponíveis no SINASC DATASUS.*

## 1. Introdução

Segundo a Organização Mundial da Saúde (OMS), Baixo Peso ao Nascer (BPN) é definido como peso inferior a 2500 gramas ao nascer independente do idade gestacional (United Nations Children's Fund 2019). Quando comparados com bebês que nasceram acima dessa linha de corte, nascidos de baixo peso possuem maior risco de óbito, podem apresentar problemas de crescimento e desenvolvimento intelectual. E quando adultos, podem apresentar maior risco de obesidade, desenvolvimento de doenças do coração e diabetes (United Nations Children's Fund 2019). Parto prematuro, má nutrição materna, consumo de álcool e cigarro durante a gestação são fatores já conhecidos que aumentam o risco de BPN.

A partir do momento que o médico possui a informação de que determinada gravidez possui alto risco de gerar um bebê com baixo peso, uma série de medidas podem ser tomadas para evitar esse desfecho. Melhorar o status nutricional da mãe e o acesso a acompanhamento pré-natal são medidas que tendem a diminuir o risco de Baixo Peso ao Nascer (United Nations Children's Fund 2019).

Além disso, existe uma série de intervenções, em diferentes níveis, para prevenir BPN. A nível nacional/regional podem ser adotados programas de proteção social, melhoria na qualidade de água, etc. Já a nível pessoa, a suplementação de nutrientes como ferro e ácido fólico contribuem para um melhor desenvolvimento da gravidez (4).

A proporção de BPN no mundo varia, em média, entre cerca de 7% e 17%, com algumas regiões chegando até 26% (United Nations Children's Fund 2019). A partir de dados obtidos no DATASUS SINASC, em 2022 a proporção de Baixo Peso ao Nascer no Brasil foi de 9%. Conseguir prever com antecedência esse risco pode ser uma importante ferramenta para auxiliar o médico na condução daquela gestação.

Trabalhos anteriores se ocuparam em desenvolver modelos de *Machine Learning* (ML) utilizando dados demográficos, condições de saúde pré-existent e informações de exames pré-natal (Ranjbar et al. 2023). Nesse artigo nós buscamos desenvolver um modelo de *Machine Learning* com o objetivo de calcular a probabilidade de Baixo Peso ao Nascer utilizando as variáveis disponíveis no Sistema de Informações sobre Nascidos Vivos (SINASC).

## 2. Revisão da Literatura

Diversos autores se ocuparam tanto em identificar fatores de risco quanto em desenvolver modelos de *machine learning* para prever baixo peso ao nascer. Chen et al. (2013), a partir de uma base de 101.163 partos em 14 diferentes regiões na China pôde identificar que idade da mãe menor que 20 anos, com baixo nível de educação e histórico de partos adversos são fatores associados ao risco de baixo peso ao nascer.

Ranjbar et al. (2023) testou o uso dos algoritmos Regressão Logística (RL), Árvore de Decisão, Floresta Aleatória, *Deep Learning Fastforward*, Light GBM, XGBoost, SVM e KNN com dados da *Iranian Maternal and Neonatal Network (IMaNet)* de janeiro de 2020 até janeiro de 2022 para desenvolver um modelo de ML capaz de prever baixo peso ao nascer. Em um universo de 8.853 partos, 1.280 (14,5%) foram de bebês com o desfecho de interesse. Além disso, os pesquisadores dispunham tanto de dados demográficos da mãe quanto sua informações de saúde e comorbidades. Para o treino e validação do modelo, a base foi separada em amostra treino (70%) e teste (30%). Já a validação interna foi realizada com *k-fold cross-validation*. Com esses dados, a equipe constatou que o XGBoost foi o melhor modelo para a tarefa, com acurácia de 0,79, precisão de 0,87, recall de 0,69 e F1 score de 0,77. E, dentre as variáveis disponíveis, idade gestacional e histórico de BPN foram os principais preditores.

Já Islam Pollob et al. (2022) trabalhou com uma base de 2351 respondentes proveniente da *Bangladesh Demographic and Health Survey* entre 2017 e 2018 para, também desenvolver um modelo capaz de prever BPN nesse contexto. Dentre as observações, 16,2% foram bebês com baixo peso ao nascer. O estudo dispunha de informações demográficas da mãe, do parto e do bebê. Os algoritmos Regressão Logística e Árvore de Decisão foram utilizados, e observou-se que a RL atingiu os melhores resultados, com acurácia de 0,87 e AUC de 0,59. No entanto, a conclusão do estudo foi de que seria necessária uma solução mais eficiente para ser utilizada na realidade de Bangladesh. Além disso, esse estudo utilizou variáveis obtidas no momento do parto, como local e tipo de parto.

Singh et al. (2013) realizou um estudo num hospital universitário de cuidados

terciários no norte da Índia, em uma base de 250 BPN e 250 com peso maior ou igual a 2.500g. Além disso, esse estudo teve acesso a informações de partos anteriores e de saúde da mãe, como ganho de peso durante a gravidez, incidência de anemia e má nutrição, nível educacional e etc. Os pesquisadores utilizaram seis variáveis preditoras, sendo elas: ganho de peso inadequado na mãe, quantidade de proteína inadequada na dieta da mãe, parto prematuro anterior, parto BPN anterior, anemia e fumante passiva. Modelando com uma regressão logística, a equipe atingiu, na base de validação, sensibilidade de 72% e especificidade de 64%.

Diferente dos trabalhos anteriores, esse estudo busca desenvolver um modelo capaz de indicar o risco de baixo peso ao nascer com base em informações pessoais e demográficas da mãe e do pai. Desse modo, na primeira consulta com o médico já seria possível entender a necessidade de cuidados especiais. Além disso, o Sistema de Informações sobre Nascidos Vivos (SINASC) possui um amplo registro dos partos no Brasil, o que nos permite trabalhar com uma base de dados muito maior do que os trabalhos anteriores.

### **3. Métodos**

#### **3.1. Base de dados**

As descobertas desse estudo foram baseadas no Sistema de Informações sobre Nascidos Vivos (SINASC). Este programa foi implementado em 1990 com o objetivo de coletar dados sobre os nascimentos em todo território nacional e para todos os níveis do Sistema de Saúde. Agrupados por ano, as bases disponibilizadas contemplam uma série de informações demográficas sobre a mãe e o pai, além de informações sobre o próprio nascido. Foram selecionados os períodos de 2022 e 2023 para comporem a base de treino e o período de 2024 para a base de teste. Até a presente data, a base correspondente a 2024 possui nascimentos registrados até o dia 11 de junho de 2024.

#### **3.2. Pré-processamento e modelagem**

A população alvo foi dividida conforme o seu peso, nascidos com menos de 2500g foram chamados Baixo Peso ao Nascer e nascidos com peso registrado maior ou igual a 2500g, Não Baixo Peso ao Nascer. Observações sem o peso registrado foram desconsiderados da base de dados. Além disso, as variáveis correspondentes a número de filhos vivos, número de gestações, número de partos normais, número de partos cesária e número de perdas fetais/abortos foram truncadas em 10, visto que uma proporção infima dos dados apresentaram valores maiores do que esse.

Para desenvolver o modelo preditor, os seguintes algoritmos foram testados na tarefa de prever Baixo Peso ao Nascer: *eXtreme Gradient Boosting* (XGB), *Light Gradient Boosting* (LGBM), *CatBoost*, *Naive Bayes* e *Random Forest*. Para selecionar quais variáveis seriam consideradas nos estudos como preditoras, duas linhas foram seguidas: variáveis sem ligação direta com o bebê, e o parto, e variáveis comuns nos estudos citados na seção de Revisão da Literatura. Desse modo, foram mantidas como possíveis variáveis dependentes a idade da mãe, idade do pai, estado civil da mãe, escolaridade em anos de estudo concluídos da mãe, número de filhos vivos e perdas fetais/abortos, número de partos normais e cesáreos, cor da mãe e tipo de gravidez (Vide Tabela 1). Todas as variáveis selecionadas possuíam valores faltantes, que foram substituídos pelo valor mais

frequente correspondente. Com exceção das variáveis ESTCIVMAE e ESCMAE, que foram substituídas por 9, visto que é um valor padrão para dados faltantes da própria base. Além disso, as variáveis categóricas foram transformadas em variáveis *dummy* segundo as suas categorias, com a técnica de *OneHotEncoder*. A linguagem de programação escolhida para desenvolver o estudo foi o Python e as bibliotecas Scikit-learn, Pandas e Seaborn foi utilizada para gerar os modelos de *Machine Learning* e as análises sobre as variáveis.

Apesar dos dados estarem desbalanceados, optou-se por não aplicar nenhuma técnica de balanceamento. Segundo (Elor and Averbuch-Elor 2022), técnicas de balanceamento de dados não melhoram a capacidade preditiva de modelos estado da arte, como os algoritmos de boosting. (Elor and Averbuch-Elor 2022) mostrou que balanceamento não melhora a capacidade preditiva de modelos Estado da Arte, como os modelos de *Boosting*. A recomendação é otimizar o *threshold* para melhorar modelos desbalanceados, o que foi aplicado neste artigo.

**Tabela 1. Variáveis preditoras**

Variável	Descrição
IDADEMAE	Idade da mãe
IDADEPAI	Idade do Pai
ESTCIVMAE	Situação conjugal da mãe: 1– Solteira; 2– Casada; 3– Viúva; 4– Separada judicialmente/divorciada; 5– União estável; 9– Ignorada
ESCMAE	Escolaridade, em anos de estudo concluídos: 1 – Nenhuma; 2 – 1 a 3 anos; 3 – 4 a 7 anos; 4 – 8 a 11 anos; 5 – 12 e mais; 9 – Ignorado
QTDFILVIVO	Número de filhos vivos
QTDFILMORT	Número de perdas fetais e abortos
QTDPARTNOR	Número de partos vaginais
QTDPARTCES	Número de partos cesáreos
GRAVIDEZ	Tipo de gravidez: 1– Única; 2– Dupla; 3– Tripla ou mais; 9– Ignorado.
RACACOR	Cor da mãe: 1– Branca, 2– Preta, 3– Amarela, 4– Parda e 5– Indígena.

A base de dados correspondente aos períodos de 2022 e 2023 foi utilizada tanto para treino quanto para validação interna, em uma proporção de 70% e 30%, respectivamente. A partir dessas amostras foram selecionados os melhores hiperparâmetros pela biblioteca Optuna, escolhidas as variáveis preditoras finais e escolhido o melhor modelo para a tarefa de prever o risco de Baixo Peso ao Nascer. Já a base de teste, dados disponíveis de 2024, foi utilizada para reportar as métricas finais de validação do modelo.

### 3.3. Métricas de validação

Já com relação as métricas de validação observadas, este trabalho reporta a matriz de confusão, acurácia, *Precision* e *Recall*. A matriz de confusão é uma ferramenta de auxílio para melhor entender a distribuição de acertos e erros do modelo. A acurácia calcula a proporção de acertos do modelo na base de dados. Dado o alto nível de desbalanceamento

da variável dependente, esta métrica deve ser observada com cuidado para não levar a interpretações incorretas da capacidade preditiva do modelo. *Precision* mede a proporção de verdadeiros positivos. *Recall* mede a assertividade dentre as observações positivas. Em outras palavras, quando a classe verdadeira era positiva, com que frequência o modelo predisse como positiva. Entendendo que o custo de falsos negativos seja maior do que o de falsos positivos para esse estudo, a métrica *Recall* foi priorizada no treino do modelo.

#### 4. Resultados

A base de treino do modelo possuía 9,5% de Baixo Peso ao Nascer, dentre 4.074.684 milhões de nascidos entre 2022 e 2023. As Tabelas 2, 4 e 3 possuem uma série de informações sobre a amostra com relação ao desfecho estudado.

A Tabela 2 mostra que as médias de idade dos pais e das mães são muito similares entre os BPN e Não-BPN. No entanto, a Tabela 3 mostra que mães com até 16 anos ou com mais de 35 anos apresentam maior proporção do desfecho. Assim como com relação a idade do pai, nos casos com até 16 anos ou maior que 45 anos.

**Tabela 2. Relação das variáveis preditoras numéricas com o desfecho BPN**

Variáveis	BPN	Não-BPN
<b>Idade da Mãe</b>		
Média	27,90	27,57
Desvio Padrão	7,21	6,68
<b>Idade do Pai</b>		
Média	32,48	32,11
Desvio Padrão	8,1	7,7

**Tabela 3. Relação das idades com o desfecho BPN**

Variáveis	BPN	% BPN
<b>Idade da Mãe</b>		
Menor ou igual 16	129.701	12,50
]16; 25]	1.549.427	9,00
]25; 35]	1.815.046	8,99
]35; 45]	575.296	11,40
Maior 45	51.75	19,70
<b>Idade do Pai</b>		
Menor ou igual 16	3.753	13,29
]16; 25]	293.143	8,71
]25; 35]	655.858	8,19
]35; 45]	383.969	9,28
Maior 45	64.477	10,43

Já na Tabela 4 pode-se observar uma maior proporção de Baixo Peso ao Nascer em mães Solteira, Viúvas e Divorciadas se comparadas com mães Casadas e em União Estável, indicando que um maior suporte na gestação tem efeito no peso do bebê. E quanto maior a escolaridade da mãe, menor a incidência de BPN. Além disso, o Tipo de Gravidez também indica alguma relação com o desfecho, visto que nos casos de gravidez

Dupla e Tripla ou Mais, a proporção de Baixo Peso ao Nascer observada foi de 62,88% e 96,05%, respectivamente.

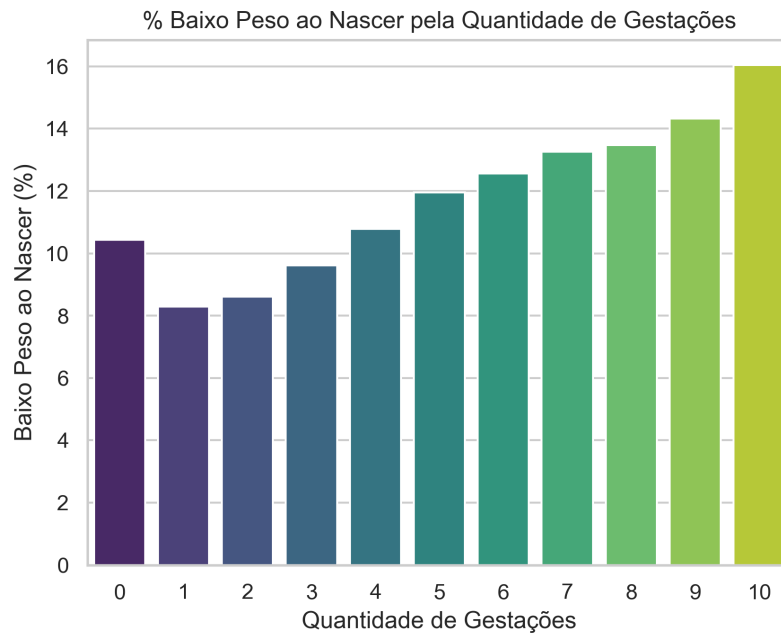
Também foi possível observar uma maior incidência de baixo peso ao nascer no caso de mães pretas. Se comparadas com mães brancas (menor valor), a proporção do desfecho é mais de 1,5% maior em mães pretas. Pode-se observar na Tabela 4 que a escolaridade da mãe tem um efeito negativo na proporção de baixo peso ao nascer: Quanto maior a escolaridade, menor a proporção do desfecho estudado.

**Tabela 4. Relação das variáveis preditoras categóricas com o desfecho BPN**

<b>Variáveis</b>	<b>BPN</b>	<b>% BPN</b>
<b>Estado Civil da Mãe</b>		
Solteira	2.040.974	10,04
Casada	1.274.511	8,81
Viúva	6.946	10,90
Divorciada	65.308	10,09
União estável	656.606	8,89
Ignorada	30.835	9,56
<b>Escolaridade da Mãe</b>		
Nenhuma	12.353	15,08
Ignorado	34.056	8,67
1 a 3 anos	43064	11,55
4 a 7 anos	431.801	10,82
8 a 11 anos	2.598.227	9,37
12 e mais	955.679	8,98
<b>Cor da Mãe</b>		
Branca	1.347.306	9,30
Preta	303.268	10,97
Amarela	18.889	9,65
Parda	2.261.900	9,33
Indígena	44.566	9,71
<b>Tipo de Gravidez</b>		
Única	3.979.454	8,21
Dupla	90.762	62,88
Tripla ou mais	1.815	96,05
Ignorado	135	31,25

A maior parte dos partos registrados são nulíparos, então vemos uma maior proporção de BPN nessa categoria. No entanto, na Figura 1 podemos observar que a proporção do desfecho aumenta conforme o número de gestações anteriores aumentam. O mesmo ocorre para a quantidade de partos normais e cesária, e para a quantidade de filhos vivos e mortos (Vide Figura 2). Para o treino do modelo a variável referente a quantidade de gestações não foi considerada, visto que ela é uma combinação das variáveis sobre a quantidade de partos normais e quantidade de partos cesáreas.

Os modelos *Light Gradient Boosting* (LGBM), *CatBoost*, *Naive Bayes* e *Random Forest* foram testados para verificar qual algoritmo desempenha melhor no problema em questão. A Tabela 5 agrupa os resultados obtidos. Todos os modelos desempenharam de



**Figura 1. Percentual de Baixo Peso ao Nascer pela Quantidade de Gestações anteriores**

forma semelhante, de modo que poderiasse seguir com qualquer um dos 4. Optou-se por utilizar o *Light Gradient Boosting* por esse tipo de modelo já ser testado na literatura para essa mesma tarefa (Ranjbar et al. 2023).

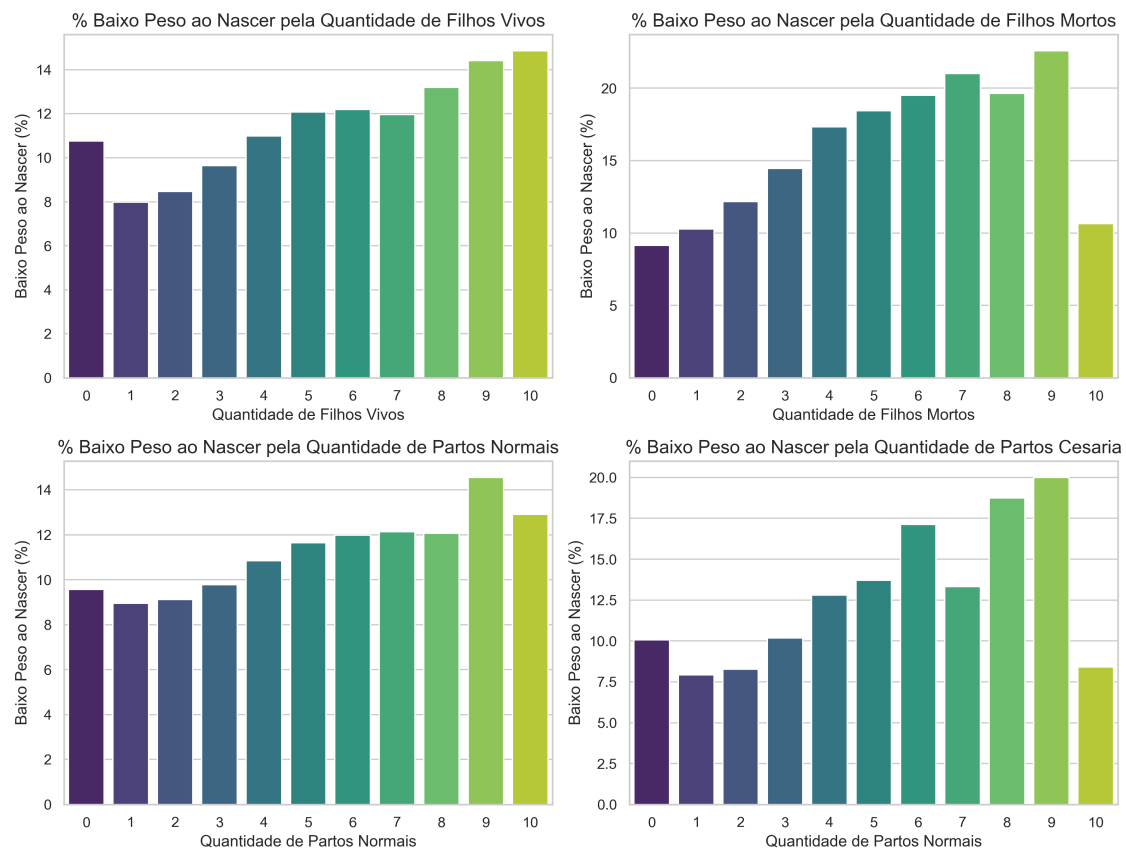
**Tabela 5. Métricas de validação para os modelos avaliados na base de validação interna**

Modelo	Recall	Precision	F1-Score	Acurácia
LGBM	0,15	0,63	0,24	0,91
CatBoost	0,14	0,64	0,23	0,91
Naive Bayes	0,15	0,64	0,23	0,91
Random Forrest	0,13	0,56	0,21	0,91

Após a escolha do modelo, aplicou-se a otimização de *threshold* com foco em melhorar o *F1-Score* na base de Validação Interna. As Tabelas 6 e 8 concentram as métricas de validação das bases de Validação Interna e Teste (jan/24 - jun/24), respectivamente. Pode-se observar que os resultados foram os mesmos em ambos os casos, indicando que não houve overfitting durante o treino do modelo e que as bases são bem homogêneas nos diferentes períodos. No entanto, os resultados obtidos não são satisfatórios.

Como o objetivo deste trabalho é prever o desfecho de Baixo Peso ao Nascer, a métrica *Recall* foi considerada a mais importante. O *Recall*, nesse caso, trata da quantidade de bebês com BPN que foram classificados como BPN pelo modelo. Pode-se observar que o *Recall* observado foi de 0,22. Ou seja, o modelo foi capaz de identificar somente 22% dos bebês registrados com baixo peso ao nascer.

No entanto, a Tabela 10 indica uma boa capacidade do modelo em prever Baixo



**Figura 2. Percentual de Baixo Peso ao Nascer pela Quantidade de Filhos Vivos/Mortos e Partos Normais/Cesáreas anteriores**

**Tabela 6. Métricas de validação do modelo na base de Validacao Interna**

Métrica	Resultado
Amostra	1.018.795
Acurácia	0,88
Recall	0,22
Precisão	0,31
F1 - Score	0,26

**Tabela 7. Matriz de Confusão na base de Validacao Interna**

	Predito Positivo	Predito Negativo
Verdadeiro Positivo	875.047	47.253
Verdadeiro Negativo	75.265	21.106

Peso ao Nascer para os bebês com maior risco. Foram criadas 4 categorias de risco de BPN com base na probabilidade do desfecho:

Desse modo, todos os nascidos no período recebem uma dessas 4 classificações conforme a sua probabilidade do desfecho. A categoria Baixo possui a maior quantidade



**Tabela 8. Métricas de validação do modelo na base de Teste**

Métrica	Resultado
Amostra	779.927
Acurácia	0,88
Recall	0,22
Precisão	0,32
F1 - Score	0,26

**Tabela 9. Matriz de Confusão na base de Teste**

	Predito Positivo	Predito Negativo
Verdadeiro Positivo	668.222	35.899
Verdadeiro Negativo	59.038	16.768

**Tabela 10. Acertos nas Faixas de Risco**

Faixa Predito	BPN	Total	Percentual de Acerto (%)
Baixo [0%-40%[	64.002	761.399	8,4
Moderado [40%-60%[	3.927	7.038	55,8
Alto [60%-80%[	7.597	11.188	67,9
Muito Alto [80%-100%[	280	302	92,7

de observações muito em função do desbalanceamento da base, e nessa categoria o modelo acerta pouco e agrupa a maior parte dos Baixo Peso ao Nascer. No entanto, conforme aumenta-se o risco de BPN, a capacidade do modelo em identificar quem realmente foi BPN aumenta. Na faixa Alto foram identificadas 11.188 observações, entre as quais 7.597 foram de bebês com baixo peso ao nascer, representando um percentual de 67,9%. Já na faixa Muito Alto, das 302 observações identificadas, 92,7% realmente foram BPN. Esses resultados nos mostram uma boa capacidade do modelo em identificar pacientes com maior risco de baixo peso ao nascer.

## 5. Discussão

A proposta do trabalho foi desenvolver um modelo de machine learning capaz de prever o risco de baixo peso ao nascer em bebês brasileiros, utilizando a base de dados do SINASC. Como se trata dos registros de nascidos vivos do Brasil, o conjunto de dados era muito grande se comparados aos utilizados em trabalhos anteriores, como (Ranjbar et al. 2023), (Islam Pollob et al. 2022) e (Singh et al. 2013). No entanto, esses trabalhos dispunham de informações não disponíveis no SINASC, como dados de saúde da mãe e do feto. Nos seus trabalhos essa variáveis desempenharam papel importante na capacidade preditiva dos modelos.

Além disso, este trabalho foi capaz de identificar fatores de risco para o desfecho de BPN e corroborar resultados encontrados anteriormente. (Chen et al. 2013) identificou que baixo nível de educação e histórico de partos adversos são fatores associados ao risco de baixo peso ao nascer. O mesmo resultado foi identificado aqui, onde a prevalência do desfecho aumenta conforme a escolaridade diminui. Também foi observado um aumento no desfecho seguindo um aumento no histórico de filhos mortos, corroborando que partos adversos aumenta o risco de baixo peso ao nascer. Outrossim, seguindo o resultado obtido

por (Ranjbar et al. 2023), o algoritmo de *Machine Learning* com melhores resultados foi o *XGBoost*.

Pôde-se observar um efeito no Estado Civil da mãe, de modo que mulheres solteiras, viúvas ou divorciadas estão mais propensas a desenvolverem gestações com BPN. Esse fator pode estar associado a falta de amparo no momento da gravidez, mas carece de estudos mais aprofundados nesse quesito. Além disso, a Cor da mãe também apresentou efeito no desfecho. Mães brancas possuem a menor incidência do desfecho, ao passo que mães pretas possuem a maior ocorrência de Baixo Peso ao Nascer.

Também foi identificado um efeito muito significativo com relação ao Tipo de Gravidez. 96,05% dos partos triplos foram BPN, o que mostra a necessidade de uma maior atenção nesse tipo de ocasião. Por fim, a Quantidade de Gestações também parece ter impacto no desfecho, de modo que quanto mais partos (normal ou cesárea) a mãe passou, maior a incidência de BPN.

Já com relação as métricas de validação do modelo, os resultados obtidos não apontam para uma boa separabilidade dos dados com as variáveis disponíveis. O *Recall* observado na base de Teste foi de 0,15, indicando que o modelo é capaz de identificar corretamente somente 15% dos nascidos com o desfecho estudado. No entanto, ao estratificar a probabilidade do desfecho por faixas de risco, pode-se observar uma boa capacidade do modelo em identificar os casos de mais alto risco, indicando uma caminho de futuras investigações e uso.

Uma vez atingido o objetivo, esse modelo poderia ser utilizado para auxiliar médicos brasileiros na condução das gestações. Mas, para tal tarefa o modelo desenvolvido carece de boa acurácia. Os resultados aqui obtidos podem servir de indicativos para médicos no momento de iniciar um acompanhamento neonatal.

## 6. Conclusão

A proposta de desenvolver um modelo de *Machine Learning* capaz de prever o risco de baixo peso ao nascer utilizando a base de dados do SINASC não foi concluído de forma satisfatória, uma vez que o modelo criado carece de capacidade preditiva no desfecho de interesse. Utilizar o DATASUS limitou o tipo de variável disponível para o desenvolvimento, indicando que somente variáveis demográficas não são suficientes para a criação de um modelo preditivo.

## 7. References

### Referências

- [Chen et al. 2013] Chen, Y., Li, G., Ruan, Y., Zou, L., Wang, X., and Zhang, W. (2013). An epidemiological survey on low birth weight infants in china and analysis of outcomes of full-term low birth weight infants. *BMC Pregnancy and Childbirth*, 13(1):242.
- [Elor and Averbuch-Elor 2022] Elor, Y. and Averbuch-Elor, H. (2022). To smote, or not to smote? *CoRR*, abs/2201.08528.
- [Islam Pollob et al. 2022] Islam Pollob, S. M. A., Abedin, M. M., Islam, M. T., Islam, M. M., and Maniruzzaman (2022). Predicting risks of low birth weight in bangladesh with machine learning. *PLoS One*, 17(5):e0267190.

- [4] Nutrition and Food Safety (NFS), W. H. O. W. N.-W. (2014). Low birth weight policy brief. Technical report, WHO.
- [Ranjbar et al. 2023] Ranjbar, A., Montazeri, F., Farashah, M. V., Mehrnoush, V., Darsareh, F., and Roozbeh, N. (2023). Machine learning-based approach for predicting low birth weight. *BMC Pregnancy Childbirth*, 23(1):803.
- [Singh et al. 2013] Singh, A., Arya, S., Chellani, H., Aggarwal, K. C., and Pandey, R. M. (2013). Prediction model for low birth weight and its validation. *Indian J Pediatr*, 81(1):24–28.
- [United Nations Children’s Fund 2019] United Nations Children’s Fund, W. H. O. (2019). Low birthweight estimates: Levels and trends 2000–2015. Technical report, WHO, Geneva.